Symbolic analysis of slow solar wind data using rank order statistics

Vinita Suyal, Awadhesh Prasad, Harinder P. Singh

Department of Physics and Astrophysics, University of Delhi, Delhi - 110007, India.

Abstract

We analyze time series data of the fluctuations of slow solar wind velocity using rank order statistics. We selected a total of 18 datasets measured by the Helios spacecraft at a distance of 0.32 AU from the sun in the inner heliosphere. The datasets correspond to the years $1975 - 1982$ and cover the end of the solar activity cycle 20 to the middle of the activity cycle 21. We first apply rank order statistics to time series from known nonlinear systems and then extend the analysis to the solar wind data. We find that the underlying dynamics governing the solar wind velocity remains almost unchanged during an activity cycle. However, during a solar activity cycle the fluctuations in the slow solar wind time series increase just before the maximum of the activity cycle.

Keywords: Solar wind plasma; Solar activity; Symbolic analysis.

1. Introduction

Solar wind is a stream of charged particles, mostly protons and electrons, that are ejected from the upper atmosphere of the sun. Parker (1958), who first coined the term 'solar wind', showed that the solar wind is in fact the outer coronal atmosphere escaping supersonically into the interstellar space. The solar wind is primarily divided into two components, a fast wind with velocities $\sim$900 km/s and a slow wind with velocities $\sim$300 km/s. These two components differ in their origin and characteristics which include flow speed, proton density, temperature, and helium content (Feldman et al., 2005; Schwenn, 2007). The fast solar wind originates from coronal holes, where the magnetic field lines are open, while slow wind is associated with open magnetic field and is highly structured and highly variable (Krieger et al., 1973; Woo and Martin, 1997).

Observations of solar wind velocity distributions made by different spacecrafts have been analyzed in considerable detail.

∗Corresponding author
Email address: hpsingh@physics.du.ac.in (Harinder P. Singh)

Helios spacecraft probed the heliosphere as close as 0.3 AU from the sun and these observations have been analyzed using linear and nonlinear time series analysis tools (Macek and Obojska, 1997, 1998; Macek, 1998; Macek and Redaelli, 2000; Redaelli and Macek, 2001; Gupta et al., 2008). Macek and Obojska (1997) argued that the inner heliosphere is a low dimensional chaotic attractor. Macek and Obojska (1998) studied a single time series using moving average, singular-value decomposition and surrogate data analysis to characterize the nonlinear behavior of the data. Macek and Redaelli (2000) have shown that the Kolmogorov entropy is positive and finite, as it holds for a chaotic system. Redaelli and Macek (2001) extended the study by using a nonlinear filter to obtain the noise-reduced data and more reliable estimate of largest Lyapunov exponent and Kolmogorov entropy. Gupta et al. (2008) analyzed several Helios time series to understand the local dynamics of fluctuations in the slow solar wind velocity. They suggest the inherent changes in the dynamics over the solar cycle using the variance in largest Lyapunov exponent for each time series.

In this work, we study how the dynamical system governing the slow solar wind velocity behaves at different phases of the solar activity cycle. We apply the method of rank order statistics of symbolic sequence (Yang et al., 2003) to investigate the solar wind time series data. In the next Section, we describe the datasets and their pre-processing. In Section 3 we describe the

Table 1: The solar wind velocity data measured by the Helios spacecraft in the years 1975 to 1981 with Carrington rotation(CR); Initial time (T_i); Final time(T_f); Data points(N). Sunspot numbers are from http://sidc.oma.be/sunspot-data/.

S.No.	CR	Year	T_i (day:h:min)	T_f (day:h:min)	N	Sunspot number
1	1625	1975	069:21:36	071:17:42	3515	18
2	1625	1975	077:00:45	078:20:52	2938	22
3	1632	1975	260:00:35	261:20:39	2979	14
4	1632	1975	267:03:52	268:23:56	3083	0
5	1639	1976	085:04:28	087:00:45	2245	38
6	1639	1976	092:07:20	094:03:10	1843	23
7	1646	1976	282:11:52	284:06:39	1188	7
8	1653	1977	099:11:25	101:07:15	2741	0
9	1653	1977	106:13:51	108:09:57	3376	31
10	1660	1977	296:17:34	298:13:48	1085	30
11	1667	1978	114:19:36	116:14:33	2393	106
12	1667	1978	121:21:18	123:17:31	1911	89
13	1681	1979	137:05:09	139:01:23	1894	148
14	1695	1980	145:10:38	147:07:04	2218	229
15	1695	1980	152:12:36	154:09:02	2559	152
16	1709	1981	159:17:43	161:14:02	3174	58
17	1709	1981	167:03:16	168:16:26	2562	119
18	1723	1982	175:00:30	176:21:15	1361	112

rank order statistics of symbolic sequence technique. In Section 4 we apply symbolic analysis technique to time series obtained from known nonlinear dynamical systems like Lorenz, Rössler, and Chua oscillators and extend the analysis to solar wind data. In the last Section we give conclusions of our study.

2. Helios data and noise reduction

We analyze time series data of radial velocity (v) measured by the Helios spacecraft at 0.32 AU in the years 1975 to 1981, corresponding to parts of the solar activity cycles 20 and 21 (June, 1976 to September, 1986). This data is obtained from
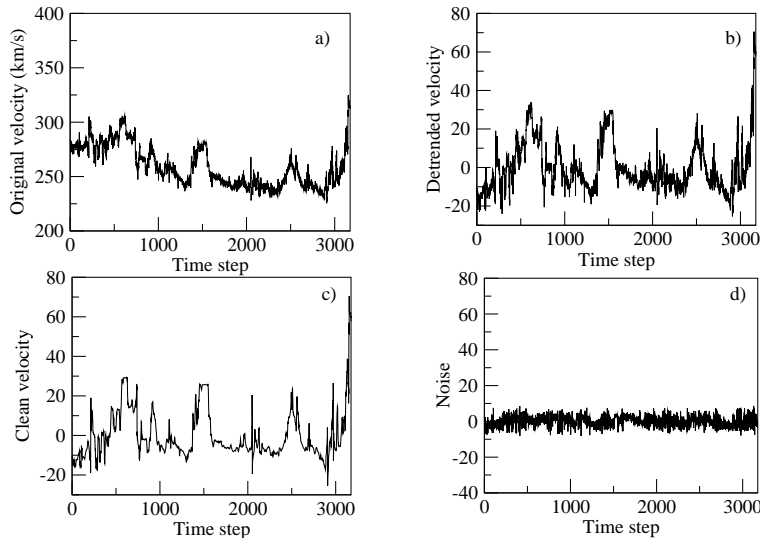
Figure 1: Pre-processing of an example data set (time series number 16 in table 1) (a) the original data, (b) the detrended data, (c) the filtered data after nonlinear noise reduction, and (d) the noise removed by nonlinear filter.

http://sprg.ssl.berkeley.edu/impact/data_browser.html using the solar wind data. Although several more time series were available, only those were selected that satisfied the condition of sufficient (> 1000) data points. Thus, 18 time series were extracted from the Helios database with the sampling time for each time series being 40.5s. The details of the time series used in the analysis are given in Table 1. In addition, Table 1 (last column) contains the corresponding sunspot numbers for each time series obtained from http://sidc.oma.be/sunspot-data/.

In Fig. 1a , we have shown a solar wind time series of 3174 data points recorded in the year 1981. The sequence shows a general decreasing trend. Therefore, before a detailed analysis is carried out, linear and quadratic trends ($v = at^2 + bt + c$, with the losing a fraction of the total sample) were subtracted from the raw data (Macek and Obojska, 1997; Gupta et al., 2008). For the data shown in Fig.1a, parameters a, b, and c are 0.15, −1.71, and 294.77 respectively. Fig. 1b shows the resulting detrended time series for the data.

Solar wind data are believed to be a output of a dynamical process upon which the noise has been superimposed (Macek and Obojska, 1998; Macek and Redaelli, 2000). Following Macek and Redaelli (2000) and Gupta et al. (2008) we use a nonlinear filter (Schreiber, 1993a) to clean the signal. This filter averages in the embedding space of a chosen dimension $2l+1$ and a defined neighborhood of size ϵ, about 2-3 times of the estimated initial noise level (Schreiber, 1993b).

3
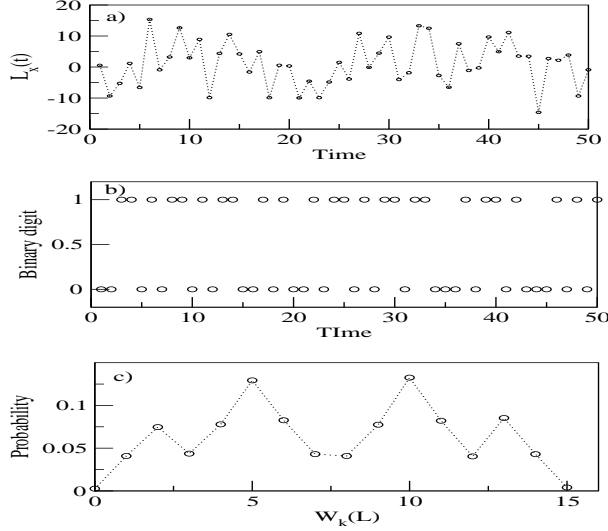
Figure 2: (a) Time series (x component) of the Lorenz oscillator, Eq. (5), (b) binary digit corresponding to the time series, binary word corresponds to $t = 1$ is $(2^3.0 + 2^2.0 + 2^1.1 + 2^0.1 = 3)$, (c) probability distribution of every 4-bit word. The word index ranges from 0 to $2^L - 1$ for L- bit words (complete time series).

Let $v_i, i = 1, 2, ..., n$ be the detrended signal. We construct vectors $V_i = (v_{i-l}, ..., v_{i+l})$ in this $2l+1$ embedding space. We then replace the data point v_i by its mean value in the neighborhood of size ϵ,

$$v_i^c = \frac{1}{N_i} \sum_{|V_i - V_j| < \epsilon} v_j, \qquad (1)$$

Where v_i^c denote the clean value and N_i is the number of neighbors within ϵ, i.e., $|V_i - V_j| < \epsilon$. This process is now iterated taking decreasing ϵ until no neighbors are found. We performed two iterations taking the parameters: $l = 3, \epsilon = 0.12$ and $l = 6, \epsilon = 0.06$ (Schreiber, 1993a). The noise (that is, the difference between the detrended data in Fig. 1b and the filtered data in Fig. 1c) is shown in Fig. 1d.

3. Symbolic Analysis

Time series that originate from dynamical systems having complex nonlinear interactions may look very much like noise (Kantz and Schreiber, 2004). Dynamical system behind such intrinsically noisy dataset can be simplified by mapping it to the binary sequences, where the increase and decrease of fluctuations are denoted by 1 and 0 (Kurths et al., 1995).

Let $x_i, i = 1, 2, ..., n$ be the time series data. A part (50 data points) of the time series (here x component of Lorenz oscillator (Lorenz, 1963)) is shown in Fig. 2a. Following Yang et al. (2003), we classify each pair of successive data into one of the two states that represents a decrease in x, or an increase in x. These two states are mapped to the symbols
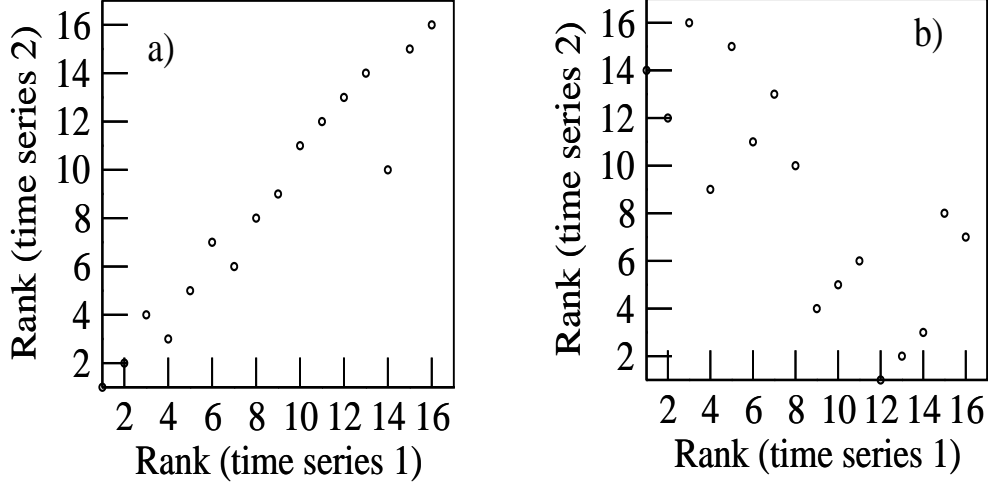
4

Figure 3: Rank order comparison of two time series from (a) same system (x component of Lorenz oscillator at different time), (b) different system (x component of Lorenz oscillator and x component of Rössler oscillator).

0 and 1, respectively:

$$S_n = 0 \quad x_n \leq x_{n-1}$$
$$= 1 \quad x_n > x_{n-1} \qquad (2)$$

We take $S_1 = 0$. This binary sequence represents a unique pattern of fluctuations in a given time series. Fig. 2b shows sequence of binary digits for the part of the data shown in Fig. 2a. The resulting symbolic series is then partitioned into short sequences of given length L. Every short sequence is then uniquely identified by a unique word (integer),

$$w_L(k) = \sum_{i=1}^{L} 2^{L-i} S_{i+k-1}. \qquad (3)$$

We find the probabilities of different words (Fig. 2c) in the complete time series, and then sort them according to decreasing frequency. The resulting rank-frequency distri-

bution, therefore, represents the statistical hierarchy of symbolic words of the time series. We compare the rank number of each L bit word for two different time series. In Fig. 3 we plot the rank number of two time series against each other. Fig. 3a shows the rank order comparison for time series of x component of chaotic solution of Lorenz oscillator at two different time regimes. In Fig. 3b rank order comparison for time series of x component of Lorenz oscillator and time series of x component of Rössler oscillator in the chaotic regime is shown. If the two time series are similar (from same dynamical system), their rank order points will be located near the diagonal line (Fig. 3a). If the two time series are from different dynamical systems, their rank points will be more scattered (Fig. 3b). We define a parameter weighted distance for L bit sequence (D_L) between the

5

two time series as:

$$D_L(x, y) = \frac{\sum_{k=1}^{2^L} \mid R_x(w_k) - R_y(w_k) \mid p_x(w_k)p_y(w_k)}{(2^L - 1)\sum_{k=1}^{2^L} p_x(w_k)p_y(w_k)}. \quad (4)$$

Here x and y are two time series, $p(w_k)$ and $R(w_k)$ represent the probability and the rank of a specific word (w_k) in the time series respectively. Value of D_L lies between 0 and 1. Greater distance indicates less similarity between the two time series and vice-versa. The distance between same two time series is zero.

4. Results and Discussion

4.1. Rank order statistical analysis of nonlinear models

We take three systems (Lorenz, Rössler, and Chua) exhibiting chaotic motions. Lorenz system is give by (Lorenz, 1963):

$$\begin{aligned}
\dot{x} &= \sigma(y - x), \\
\dot{y} &= x(\rho - z) - y, \\
\dot{z} &= xy - \beta z. \quad (5)
\end{aligned}$$

We choose parameters $\sigma = 10, \rho = 28$, and $\beta = 8/3$.
Rössler system is given by (Rössler, 1976):

$$\begin{aligned}
\dot{x} &= -y - z, \\
\dot{y} &= x + ay, \\
\dot{z} &= b + z(x - c). \quad (6)
\end{aligned}$$

Parameters are fixed at $a = 0.1, b = 0.1$, and $c = 18$.

Chua system is given by (Chua et al., 1993):

$$\begin{aligned}
\dot{x} &= c_1(y - x - p(x)), \\
\dot{y} &= c_2(x - y + z), \\
\dot{z} &= -c_3 y, \quad (7)
\end{aligned}$$

where $p(x)$ is defined as, $p(x) = m_1 x + ((m_0 - m_1)(\mid x + 1 \mid - \mid x - 1 \mid))/2$. We take $c_1 = 15.6$, $c_2 = 1, m_0 = -8/7$, $m_1 = 5/7$ and select the parameter $c_3 = 33$ such that the motion will be a single-scroll type.

We take two time series from each system at different time intervals. Hence we have 6 time series in all. They are denoted as $L_x(t)$ and $L_x(t')$ for the Lorenz oscillator, $R_x(t)$, $R_x(t')$ for the Rössler oscillator, and $C_x(t)$ and $C_x(t')$ for the Chua Oscillator. We use rank order statistics on these time series data by keeping $L = 4$. Fig. 4a shows distances (D_L) for different combinations of these time series of length 1000, 5000 and 10000. It is clear from Fig. 4a that time series from the same system have less distance as compared to time series from different systems. Fig. 4a also indicates that 1000 data points are sufficient for this analysis. Similar results are obtained when we add noise to the original time series. Fig. 4b shows the result when we add 5%, 10%, and 20% noise to the original time series of 1000 data points. Therefore, we can say that this rank order statistics is quite robust against the presence of high level external noise.

4.2. Rank order statistical analysis of solar wind

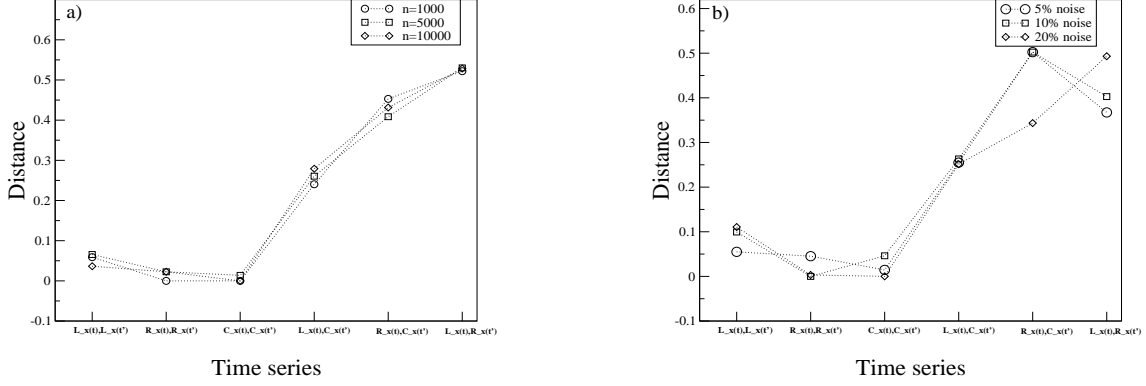Solar wind data have positive largest Lyapunov exponent (Pavlos et al., 1992;

6

Figure 4: Distances D_L for different combinations of time series obtained from nonlinear dynamical systems (a) for time series of length 1000, 5000, and 10000, (b) with 5%, 10%, and 20% noise added to the time series of 1000 data points.

Macek and Redaelli, 2001; Gupta et al., 2008), positive Kolmogorov entropy (Macek and Redaelli, 2000, 2001), and fractal dimension (Macek and Obojska, 1997). This leads to the conclusion that wind dynamics are driven by the complex nonlinear interactions. We use rank order statistics with $L = 4$ and time series of 1085 data points for the analysis. Fig. 5 shows the solar activity from March 1975 to September 1986. It shows monthly averaged sunspot index and smoothed monthly sunspot index. In the inset daily sunspot index around the maximum of solar activity cycle is shown. We have shown starting time of our analyzed time series on the activity cycle. Time series 13, 14, and 15 are located around the maximum of solar activity cycle 21. Sunspot index corresponding to them are 148, 229, and 152 respectively i.e., sunspot index increases while going from time series 13 to 14 and decreases on going from time series 14 to 15.

Fig. 6 shows the inter-time series distance of solar wind velocity as given by Eq. (4). Behavior of inter-time series distance remains same for all the time series, but the distance is maximum with time series 13 (for time series $1, 2, 3, 7, 9, 10, 11, 12, 14, 15, 18$) and with time series 14 (for time series $4, 5, 6, 8, 13, 16, 17$), i.e., only two time series 13 and 14 show least correlation with rest of the time series. Fig. 7 shows the most distant time series from a particular time series. Positions of these time series are around the maximum of the solar activity cycle but towards the increasing phase only. Time series 15 is also near the solar maximum but it is in the decreasing phase. This could be a reason for its less distance to the other time series as compared to the time series 14.

5. Conclusions

We used rank order statistics to analyze the time series obtained from known nonlinear systems. This method is an effective tool
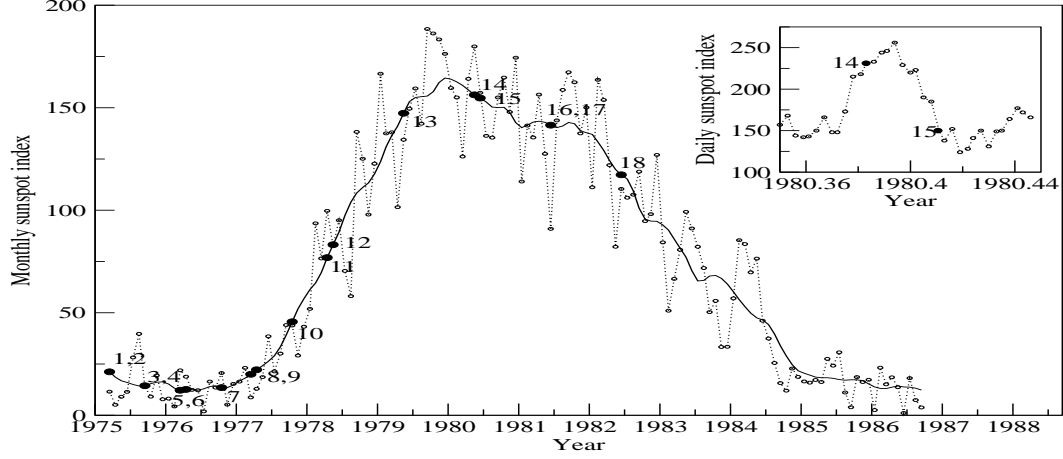
Figure 5: Monthly averaged sunspot number (broken line) and smoothed monthly sunspot number (solid line) from March 1976 to September 1986. Initial positions of solar wind data sets are shown on the smooth curve. In the inset daily sunspot index and position of data sets 14 and 15 are shown.
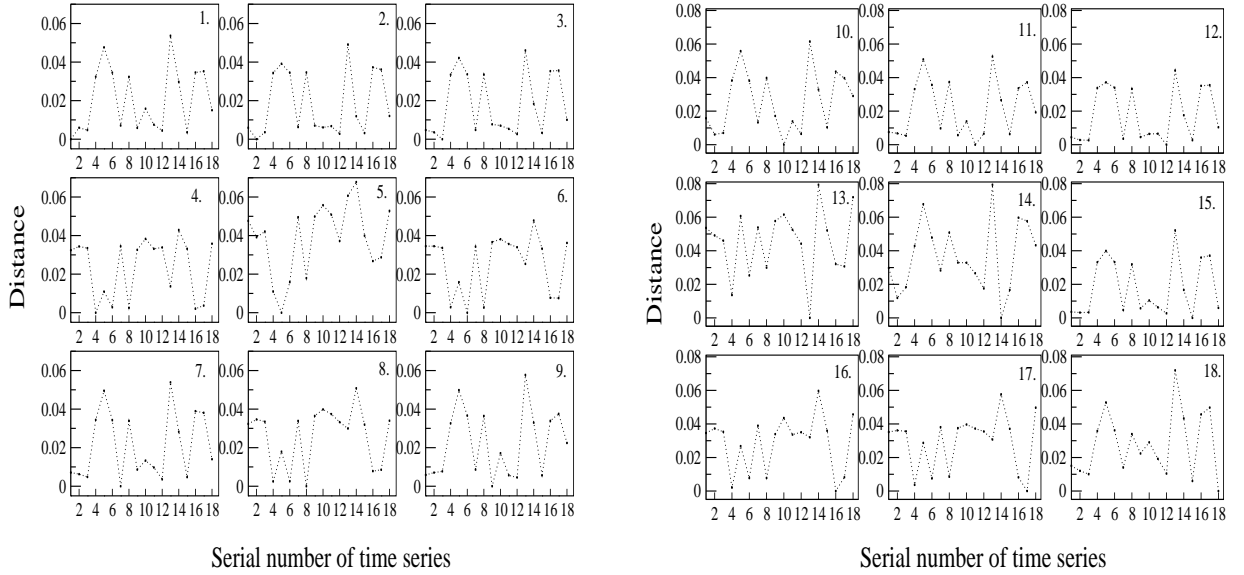

Figure 6: Distances D_L of a time series from others. For example, panel 1 (top left) shows the distance between time series 1 and all 18 time series. The distance between same two time series is zero.
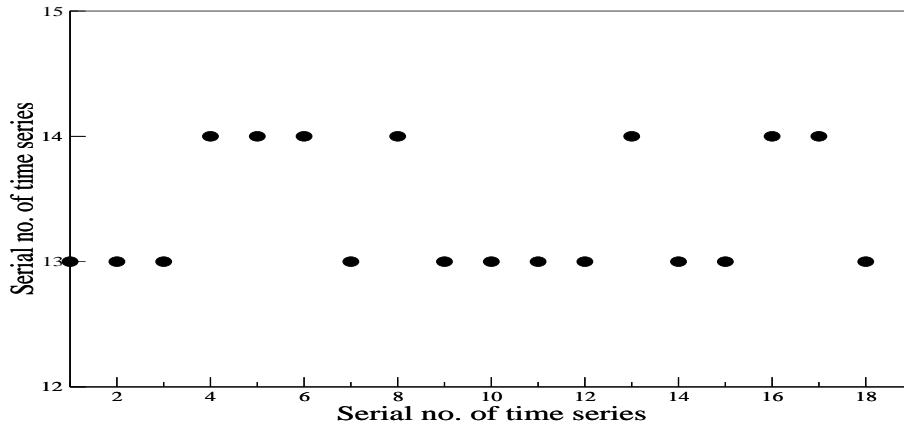
8

Figure 7: Serial number of time series vs. most distant time series serial number.

for analyzing short time series even in the presence of high level of external noise and is applied to 18 solar wind time series obtained by the Helios spacecraft for the years $1975 - 1982$. Nonlinear filter is used to reduce the noise present in the data. We found that all the time series have least correlation with data sets 13 and 14, which are located near the maximum of the solar cycle. Although data set 15 is also located very close to the data set 14 (on the solar activity cycle), it shows significant correlation. One possible reason for this may be that data set 15 is located in the decreasing phase of the solar activity cycle, while data sets 13 and 14 are located in the increasing phase of the solar activity cycle. This suggests the presence of larger fluctuations in the solar wind velocity data corresponding to the time just before the maximum of the activity cycle.

References

Chua, L. O., Itoh, M., Kocarev, L., Eckert, K., 1993. J. Circ. Syst. Comput. 3, 93.

Feldman, U., Landi, E., Schwadron, N. A., 2005. J. Geophys. Res. 110, 07109.

Gupta, K., Prasad, A., Saikia, I., Singh, H. P., 2008. Planetary and Space Science 56, 530.

Kantz, H., Schreiber, T., 2004. Nonlinear Time Series Analysis. Cambridge University Press.

Krieger, A. S., Timothy, A. F., Roelof, E. C., 1973. Solar Physics 29, 505.

Kurths, J., Voss, A., Saparin, P., Witt, A., Kleiner, H. J., Wesse, N., 1995. Chaos 5, 88.

Lorenz, E. N., 1963. Journal of the Atmospheric Sciences 20, 130.

Macek, W. M., 1998. Physica D: Nonlinear Phenomena 122, 254.

Macek, W. M., Obojska, L., 1997. Chaos, Solitons & Fractals 8, 1601.

Macek, W. M., Obojska, L., 1998. Chaos, Solitons & Fractals 9, 221.

Macek, W. M., Redaelli, S., 2000. Phys. Rev. E 62, 6496.

Macek, W. M., Redaelli, S., 2001. Advances in Space Research 28, 775.

Parker, E. N., 1958. Astrophys.J. 128, 664.

Pavlos, G. P., Kyriakou, G. A., Rigas, A. G., Liatsis, P. I., Trochoutsos, P. C., Tsonis, A. A., 1992. Annales Geophysicae 10, 309.

Rössler, O. E., 1976. Phys. Lett. A 57, 397.

Redaelli, S., Macek, W., 2001. Planet. Space sci. 49, 1211.

Schreiber, T., 1993a. Phys. Rev. E 47, 2401.

Schreiber, T., 1993b. Phys. Rev. E 48, 1752.

Schwenn, R., 2007. In: Baker, D. N., Klecker, B., Schwartz, S. J., Schwenn, R., von Steiger, R. (eds.), Solar Dynamics and Its Effects on the Heliosphere and Earth. Vol. 22 of Space Sciences Series of ISSI. Springer New York, p. 51.

Woo, R., Martin, J. M., 1997. Vol. 20. p. 2535.

Yang, A. C.-C., Hseu, S.-S., Yien, H.-W., Goldberger, A. L., Peng, C.-K., 2003. Phys. Rev. Lett. 90, 108103.